

EPA'S RESPONSE TO LWG AUGUST 26, 2014 REQUEST FOR DISPUTE RESOLUTION

A. INTRODUCTION

The LWG objects to EPA's approach to defining background based on upriver bedded sediment data. They argue EPA's approach is wrong in two ways: (1) analytically valid data is discarded as outliers contrary to EPA guidance, and (2) numerous statistical errors were made in determining which data to exclude as outliers. The LWG further argues that even if EPA was consistent with its guidance, that national guidance is flawed and should not be relied upon. Additionally, they assert that even if their alleged statistical errors were ignored, the outlier data means that elevated concentrations exist upstream and may be transported downstream to the Portland Harbor Site which, in their view, necessarily means the sediment cleanup will be recontaminated above EPA's calculated background concentrations. Last, the LWG argues that EPA has abused its discretion in its approach to determining background in Section 7. EPA disagrees with the LWG's arguments and provides its specific responses in more detail below.

The LWG and EPA have disagreed on a background evaluation approach for a long time, since at least 2008 when EPA commented on the LWG's background discussion in its Comprehensive Round 2 Site Characterization Summary and Data Analysis Report. There is a long record of comments, technical memoranda, and discussions on background. (See Exhibit 1 for a summary of the many comments on background). In all of this time, the LWG has consistently maintained that the data set contained no outliers and that none should be eliminated from the data set. Ultimately, EPA provided specific direction to the LWG to eliminate outliers in the draft RI report. The LWG could have raised a dispute at that time after it received specific direction from EPA. Instead, the LWG chose to provide a draft document that didn't follow EPA direction and this is one of the reasons why EPA is now revising the RI report. This longstanding disagreement on background is now being formally disputed and brought to the ECL Office Director for resolution.

EPA's background evaluation approach and its calculation of background concentrations for the indicator contaminants as set forth in the EPA-modified Section 7 of the RI Report are consistent with the NCP and relevant EPA guidance. Additionally, the LWG failed to provide any information beyond the background data itself to support its argument that the sediment represented by the outlier samples is of such volume or mass that it will significantly impact overall sediment quality at the Portland Harbor site if transported to the Site.

Our responses below are organized to follow the specific issues raised by the LWG on page 2 in their letter of August 26, 2014. However, abuse of discretion is a common theme in some LWG arguments on technical issues. EPA is responding to this issue and providing the appropriate standard of review for CERCLA decisions separately from the technical issues.

B. SUMMARY OF EPA'S REQUESTED DECISION FROM THE DISPUTE OFFICIAL

EPA requests the dispute official to determine that:

1. EPA did not abuse its discretion and EPA's modified Section 7 of the RI Report is acceptable and consistent with EPA guidance;

2. The LWG must incorporate the modified Section 7¹ into the final draft of the RI Report;
3. The LWG must incorporate the information provided in Tables 1 through 6 in EPA's discussion in [E], below, as the tables supporting Section 7 discussion in the final draft of the RI Report²;
4. EPA's approach to calculating background concentrations from the background data set must be used by the LWG to calculate background for the remaining 23 contaminants needing such analysis for the final RI Report. In order to assure that the calculation of background on the remaining 23 contaminants is in accordance with this decision and Section 7 modifications, the dispute official should request the LWG to provide its calculations for the 23 contaminants to EPA's project manager no later than 30 days from the date of the final decision for EPA review and approval; and
5. His decision is based on the Administrative Record for this dispute.

C. STANDARD OF REVIEW AND LWG ISSUE OF ABUSE OF DISCRETION

The LWG has framed many of their concerns about EPA's technical approach as "an abuse of discretion" on EPA's part. Abuse of discretion is a judicial review standard for challenging agency decisions under the Administrative Procedures Act, but not one CERCLA specifies as relevant to CERCLA decisions. CERCLA provides that challenges to remedy decisions shall be limited to the administrative record and the standard of review is arbitrary and capricious or otherwise not in accordance with law. 42 U.S.C. §9613(j)(1-2). Since this dispute concerns an important topic for the selection of the remedy for the Portland Harbor Superfund site, it is important that the correct standard of review be applied to this decision and is based on the written record developed during the dispute process.

The arbitrary and capricious standard is a high bar requiring a clear error of judgment or errors of procedure or glaring omissions or mistakes on the agency's part. *Motor Vehicle Manufacturers Association v. State Farm Mutual*, 463 U.S. 29, 43 (1983); *United States v. Akzo Coatings of America, Inc.*, 949 F.2d 1424, 1426. The Ninth Circuit only finds a decision arbitrary or capricious if "the agency has relied on factors which Congress has not intended it to consider, entirely failed to consider an important aspect of the problem, offered an explanation for its decision that runs counter to the evidence before the agency, or is so implausible that it could not be ascribed to a difference in view or the product of agency expertise." *Dioxin/Organochlorine Ctr. v. Clarke*, 57 F.3d 1517, at 1521 (9th Cir. 1995) (quoting *Motor Vehicle Mfr. Ass'n v. State Farm Ins.*, 463 U.S. 29, 44 (1983)). The Court's deference to the agency "is highest when reviewing an agency's technical analyses and judgments involving the evaluation of complex scientific data within the agency's technical expertise." *League of Wilderness Defenders Blue Mountains Biodiversity Project v. Allen*, 615 F.3d 1122, 1130 (9th Cir. 2011). "Agencies have discretion to rely on their own experts' reasonable opinions to resolve a conflict between or

¹ Exhibit 2 to this Response is the text of the modified Section 7.

² The original direction to the LWG requested them to develop the Tables and Figures based on EPA's worksheets, but since EPA needed to produce them for this dispute, the LWG should now be directed to incorporate them in the final RI Report.

among specialists, even if [a reviewing court] find[s] contrary views more persuasive. *Greater Yellowstone Coal. v. Lewis*, 628 F.3d 1143, 1148 (9th Cir. 2010).

Courts tend to use the abuse of discretion review in situations where a degree of discretionary decision-making has been granted to the agency, and the question is whether the agency has surpassed the intended scope of that discretionary decision-making. *See N.L.R.B. v. French Intern. Corp.*, 999 F.2d 1409 (9th Cir. 1993) (holding that to avoid an abuse of discretion an agency need only make a discretionary choice “within the range of [choices] appropriate under the circumstances.”) Courts therefore often determine first the range of discretion the governing statute affords the agency to take the challenged action, and then determine whether the action fits within that permissible range. *See Rybachek v. U.S. E.P.A.*, 904 F.2d 1276, 1297 (9th Cir. 1990) (holding that the CWA leaves EPA “with broad discretion in deciding how non-water quality environmental impacts will be taken into account.” The court then concluded that EPA’s limited consideration of the non-water quality environmental impacts fit within their broad discretion and therefore was not an abuse of discretion.)

Although factual determinations are primarily governed by the arbitrary and capricious standard of review in one case the 9th Circuit applied the abuse of discretion standard to the circumstance where evidence supporting the agency’s decision was so slight and so thoroughly outweighed by contrary evidence, that it was held to be an abuse of discretion to base a decision on such little evidence. *See Lauvik v. I.N.S.*, 910 F.2d 658, 660 (9th Cir. 1990).

As this Response and supporting information will show, EPA’s approach for evaluating the background data set was well within the discretionary decision-making authority of EPA and consistent with national agency guidance related to statistical approaches for analyzing data. The LWG has not provided any creditable basis for discrediting the use of EPA’s national guidance at the Portland Harbor Superfund Site. No clearly erroneous judgments or omissions have been made in assessing background for the Portland Harbor Superfund Site; therefore, EPA’s approach is not arbitrary and capricious.

D. EPA GUIDANCE RELEVANT TO BACKGROUND AND THE PORTLAND HARBOR SUPERFUND SITE

EPA’s Contaminated Sediment Guidance (USEPA 2005, Section 2.1.3) states that “Where site contaminants may also have natural or anthropogenic (man-made) non-site-related sources, it may be important to establish background or reference data for a site.” EPA policy on the role of background in the CERCLA cleanup program is presented in the *Role of Background in the CERCLA Cleanup Program* (USEPA 2002b). To assist Regions with developing a reliable representation of background for CERCLA sites, EPA developed a peer-reviewed practical guide to sampling and statistical analysis of background concentrations in soil at CERCLA sites (USEPA 2002a) and developed software that can be used to perform rigorous statistical analyses to help decision makers and project teams in making correct decisions at a polluted site that are cost-effective, and protective of human health and the environment (USEPA 2009a, USEPA 2013a and USEPA 2013b).

“Background” refers to constituents or locations that are not influenced by the releases from a site, and is usually described as naturally occurring or anthropogenic (USEPA 2002a, p.1-2; USEPA 2002b, page 5):

- 1) Anthropogenic – natural and human-made substances present in the environment as a result of human activities (not specifically related to the CERCLA release(s) in question); and,
- 2) Naturally occurring – substances present in the environment in forms that have not been influenced by human activity.

EPA guidance defines a defensible background data set as representing a single population possibly without any outliers. [USEPA 2013b, p.16] However, the guidance is also clear that: “[i]n a background data set, in addition to reporting and/or laboratory errors, statistical outliers may also be present. A few elevated statistical outliers present in a background data set may actually represent potentially contaminated locations belonging to impacted site areas and/or possibly from other polluted site(s); concentrations represented by those elevated outliers may not be coming from the main dominant background population under evaluation. Since the presence of outliers in a data set tends to yield distorted (incorrect and misleading) values of the decision making statistics (e.g., UCLs, UPLs and UTLs), outliers **should not be included** in background data sets and estimation of BTVs.” [(emphasis added) USEPA 2013b, p.16]. EPA guidance further provides that “[t]he objective is to compute background statistics based upon the majority of the data set representing the main dominant background population, and not to accommodate a few low probability high outliers coming from extreme tails of the data distribution that may also be present in the background data set. The occurrence of elevated outliers is common when background samples are collected from various onsite areas. The proper disposition of outliers, to include or not include them in statistical computations, should be decided by the project team. The project team may want to compute decision statistics with and without the outliers to evaluate the influence of outliers on the decision making statistics.” [USEPA 2013b, p.17]

The Portland Harbor Site is located in an industrialized harbor with a large number of historical and ongoing point and non-point sources from many potentially responsible parties. Likewise, Portland Harbor is the end of a large watershed with a large number and nature of point and non-point discharges to the river. At Portland Harbor, industrial facility sources of releases to the site have been identified as far upstream as River Mile 14, e.g., the Zidell Marine Corporation facility at Moody Avenue. Likewise, other sources in the downtown reach have been identified as sources to the Superfund Site, such as PGE at RM 13 east, and the Former Portland Gas Manufacturing at RM 12.4 west. Therefore, the data set used for calculating background for the Portland Harbor Site was taken from sediments further upriver (RM 15.3 to 28.4) above river mile 14.

The most relevant policy and guidance documents that were used by EPA to modify Section 7 are:

- *Determination of Background Concentrations of Inorganics in Soils and Sediments at Hazardous Waste Sites.* (USEPA 1995)
- *Role of Background in the CERCLA Cleanup Program.* (USEPA 2002b)

- *Guidance for Comparing Background and Chemical Concentrations in Soil for CERCLA Sites.* (USEPA 2002a)³
- *ProUCL Version 5.0.00 Users Guide.* (USEPA 2013a)
- *ProUCL Version 5.0.00 Technical Guide.* (USEPA 2013b)
- *Scout 2008 Version 1.0 User Guide (Second Edition, December 2008).* (USEPA 2009a)

The LWG relies on the Statistical Analysis of Groundwater Monitoring Data at RCRA Facilities, Unified Guidance (USEPA 2009b) in several arguments throughout their statement. EPA does not agree that the use and application of USEPA 2009b is relevant and appropriate to evaluating background for the Portland Harbor Site. The intended purpose of this guidance (p. iii) is cited in document as:

“The guidance is primarily oriented towards the groundwater monitoring statistical analysis provisions of 40 CFR Parts 264.90 to .100. Similar requirements for groundwater monitoring at solid waste landfill facilities under 40 CFR Part 258 are also addressed. These regulations govern the detection, characterization and response to releases from regulated units into the uppermost aquifer. Some of the methods and strategies set out in this guidance may also be appropriate for analysis of groundwater monitoring data from solid waste management units subject to 40 CFR 264.101. Although the focus of this guidance is to address the RCRA regulations, it can be used by the CERCLA program and for improving remedial actions at other groundwater monitoring programs.”

Whereas the guidances EPA has relied on were specifically developed in part to assess sediments (as noted in USEPA 2005) and are more relevant for this site.⁴

E. EPA’S EVALUATION OF THE BACKGROUND SEDIMENT DATA SET

EPA’s modifications to Section 7 of the RI clearly explain the process used to develop background sediment concentrations; however, because the LWG’s arguments confuse or inaccurately describe what EPA did, we will augment the modified Section 7 discussion here with additional background details and relevant guidance references.

The text of Section 7 discusses only a subset of contaminants of concern, termed “indicator contaminants,” because the main text of the RI focuses on the indicator contaminants, as agreed to by EPA and the LWG, while additional contaminants would be addressed in appendices to the Report as appropriate. Therefore, this Response discusses EPA’s evaluation of the background data set for the indicator contaminants only. However, the LWG has been directed to calculate background statistics for 23 additional contaminants, consistent with the modified Section 7 in finalizing the RI Report. Indicator contaminants are aldrin, arsenic, chlordane, chromium, copper, DDx, dieldrin, bis(2-ethylhexyl)phthalate, mercury, total PAHs, total PCBs as Aroclors, total PCBs as congeners, total PCDFs/PCDDs, tributyltin, and zinc.

³ Although this guidance is written specifically for soil, many of the concepts may be applicable to contaminant data for sediment and biota (USEPA 2005, p. 2-6)

⁴ We note the LWG had not relied on the Unified Guidance in its previous drafts of the RI Section 7.

As stated in section 7.2.2, the upriver reach of the Lower Willamette River extending from RM 15.3 to 28.4 was selected by the LWG, in consultation with EPA, DEQ, and the tribes, as the reference area for determining background sediment concentrations. This area was chosen because it is considered broadly representative of the upstream sediment concentrations resulting from point and non-point sources within the broader watershed and unaffected by either localized point sources or Portland Harbor itself. The samples were collected with the intent that the grain size distribution in the background area samples would approximate the grain size distribution of sediments the Portland Harbor Study Area.⁵

EPA used the upriver reach data sets developed by the LWG, as discussed in Sections 7.2.3 and 7.2.5. The data included summed values for total PCBs as Aroclors, total PCBs as congeners, total dioxins/furans, total PAHs, and total chlordanes. Total Aroclors, congeners, dioxins/furans, chlordanes, and PAHs were calculated using the rules described in Section 7.2.5.

Consistent with EPA statistical guidance (USEPA 2013a, b),⁶ EPA reviewed the LWG-generated figures for indicator contaminants in their 2011 draft final RI in conjunction with a review of the data files. [Exhibit 3 presents the LWG figures with EPA's visual and observational markings] These figures include scatter plots of results by river mile, normal Q-Q plots, and box-whisker plots. Contrary to the LWG's representation, EPA did review the graphical techniques prepared by the LWG. The scatter plots were utilized to observe the concentration range of reported values as well as the location(s) of those values that seemed unrelated to the dominant range of values. Normal Q-Q plots were observed to ascertain whether the data appeared to fit a single, continuous population, and whether values at the upper end of the distribution appeared discordant from the dominant population. As directed by EPA the LWG evaluated the presence of potential outliers in the 2011 draft RI (see red data points identified by LWG and EPA notes in Exhibit 3). Aldrin, dieldrin, and tributyltin were not further evaluated due to either insufficient number of detected values or insufficient number of data points.

EPA's review of the scatter plots revealed that in most instances, the majority of the data fall within a well-defined concentration range. The scatter plots were also useful for observing the degree of separation of the highest values from the majority of the data, and whether those highest values exhibited any discernable spatial distribution. The figures were reviewed in conjunction with the contaminant-specific data, and this review revealed two notable data issues

⁵ Risk-based cleanup goals are calculated as dry-weight concentrations. The LWG developed a method to OC-correct the background statistics, and EPA accepted this methodology to deal with the difference between TOC in the background data set with TOC within the study area.

⁶ USEPA 2013a, at p. 74 suggests that the outlier identification procedures be supplemented with graphical displays such as normal Q-Q plots and box plots. On a normal Q-Q plot, observations that are well separated from the bulk (central part) of the data typically represent potential outliers needing further investigation. Also, significant and obvious jumps and breaks in a normal Q-Q plot are indications of the presence of more than one population. Data sets exhibiting such behavior of Q-Q plots should be partitioned out into component sub-populations before estimating EPC terms or BTVs.

that had not been previously discussed.⁷ A summary of the resulting data sets is provided in Table 1.

Table 1. Summary of Upriver Reach Data Sets

Contaminant	No. Samples	Detects	# Non-detects
Aldrin	48	6	42
Arsenic	71	71	0
BEHP	67	52	15
Chromium	65	65	0
Copper	67	67	0
Chlordanes	48	33	15
Dieldrin	47	7	40
DDx	48	47	1
Mercury	61	52	9
PAHs	71	60	11
PCBs as total Aroclors	48	23	25
PCBs as total congeners	33	33	0
total PCDD/Fs	33	33	0
Tributyltin ion	3	2	1
Zinc	67	67	0

Given the revisions made to the data sets described in footnote 7, EPA conducted a statistical analysis of the background data. As recommended by USEPA 2013b, EPA conducted this statistical analysis on the full revised upriver reach data set for each contaminant (Table 1). EPA used ProUCL 5.0 for goodness-of-fit (GOF) tests and to calculate the background threshold value (BTv) and upper confidence level (UCL) statistics using a variety of statistical methods.⁸ For data sets with non-detects (NDs), ProUCL has several estimation methods including the Kaplan-Meier (KM) method, regression on order statistics (ROS) methods, and substitution methods (e.g., replacing NDs by DL, DL/2). EPA didn't select one method to determine the distribution and compute statistics, but let ProUCL identify how well the data fit various distributions. EPA did not select any background statistics for Section 7 of the RI Report based on the substitution methods provided in ProUCL. This is shown in the spreadsheet EPA provided

⁷ First, the presence of consistently non-detect observations of organochlorine contaminants (PCBs as Aroclors, DDx, chlordanes, aldrin, and dieldrin) above RM 23, and that the reported detection limits were consistently greater than either the detection limits or detected values of the Aroclor and chlordane data collected downstream of RM 23. Because the detection limits of the SOM01.2 analyses were consistently higher than the SW8082 results, EPA determined that the two data sets were not sufficiently similar to combine, and that it was appropriate to exclude the SOM02.1 data from the background data set. The LWG agreed to this in negotiations with EPA on November 26, 2013. Second, a comparison of the PCB results reported as Aroclors and as congeners indicated a lack of comparability between paired results at the low concentrations observed in the background data set. Consequently, EPA determined that it was inappropriate to calculate a combined "total PCBs" background as had been done previously by the LWG, and calculated separate background values for PCBs analyzed as Aroclors and as congeners. The LWG also agreed to this in negotiations with EPA on November 26, 2013.

⁸ EPA calculated the appropriate statistical endpoints on the background data set using ProUCL 5.0. [USEPA 2013a, p. 19]

- A 95% UPL was computed to represent a not-to-exceed upper threshold concentration as a pre-specified cleanup standard. EPA ran the goodness of fit tests on both the full data sets and the data sets with outliers removed.
- A 95% UCL was computed to represent an average concentration as a pre-specified cleanup standard. (Represents average exposure) EPA calculated UCLs for the full data sets and the data sets with outliers removed.

EPA appropriately calculated the statistical endpoints using ProUCL 5.0 and did not use DL/2 to conduct this evaluation. [USEPA 2013a, p. 27]

to the LWG on August 2, 2013, tab “revised tables” [Exhibit 4] and presented in Table 2⁹. Chromium, copper, and mercury were not further analyzed because no outliers were identified and the full data sets were normally distributed.

Table 2. Results of EPA Evaluation on Full Upstream Data Set (corrected)

Contaminant	Distribution(s)	Units	BTV (95% UPL)	Calculation Method	CT (95% UCL)	Calculation Method
Arsenic	Gamma	mg/kg	3.98	Gamma-HW	3	Gamma
BEHP ¹	Gamma	µg/kg	167	Gamma-HW	52	Gamma-KM
Chromium	Normal	mg/kg	32.2	Normal-t	23.8	Normal-t
Copper	Normal	mg/kg	37.4	Normal-t	25.9	Normal-t
Chlordanes	Gamma, Lognormal	µg/kg	0.91	Gamma-HW	0.43	Gamma ROS
DDx	Normal	µg/kg	4	Normal KM-t	2.3	Normal KM-t
Mercury	Normal	mg/kg	0.052	Normal KM-t	0.034	Normal-t
PAHs	Gamma	µg/kg	353	Gamma-HW	106	Gamma ROS
PCB as total Aroclors	None	µg/kg	57.3	KM-Chebyshev	10.5	KM-t
PCBs total congener	Lognormal	µg/kg	20.5	Lognormal-t	9.3	Lognormal-H
Total PCDD/Fs total	Gamma	µg/kg	0.24	Gamma-HW	0.094	Gamma
Zinc	Normal	mg/kg	110.5	Normal-t	79	Normal-t

Note 1: EPA evaluated descriptive statistics without the obvious outlier value of 2,100 µg/kg.

As noted above, EPA reviewed the scatter and Q-Q plots for the presence of suspected data outliers. EPA also conducted a statistical evaluation of outliers on the full data set utilizing Rosner’s test, which is consistent with the recommendations in ProUCL, and is applicable when the sample size has a minimum of 25 results. EPA set the maximum number of outliers to 10¹⁰, non-detect results were set at DL/2¹¹. The number of outliers identified by these tests is provided in Table 3. Although it is not necessary for the data to be normally distributed to apply Rosner’s test, the resulting data after the potential outliers are removed should follow a normal distribution. However, this condition was not met in all instances, and in such instances emphasis was given to the visual examination of the data to supplant the results of the statistical tests. This is consistent with EPA guidance for using ProUCL (USEPA 2013b, p. 188,189).

⁹ In a meeting with the LWG on November 26, 2013, it was noted that there were some errors in the evaluation that were subsequently corrected. Those corrections are reflected in this table.

¹⁰ Rosner’s test can be used to identify up to 10 outliers in data sets of sizes 25 and higher. EPA set the maximum number of outliers to be identified at 10; this did not require the program to find 10 outliers, but rather limited the program from finding more than 10 outliers. EPA did this so as not to bias the evaluation to observations made by the reviewer, but to confirm the number of outliers observed matched the number identified by the statistical test. In no instance were 10 outliers identified for any contaminant data set (see Table 3).

¹¹ For data sets with NDs, two options are available in ProUCL to deal with data sets when conducting Dixon’s and Rosner’s tests. These options are: 1) exclude NDs and 2) replace NDs by DL/2 values. These options are used only to identify outliers and not to compute any estimates and limits used in decision-making process. [USEPA 2013a, p.74]

Table 3: Results of Outlier Analysis

Contaminant	No. Outliers Observed	Notes	LWG Figure 2011 Draft RI	No. Statistically Significant Outliers	Distribution with Outliers Removed
Aldrin	NA	Insufficient detections	7.3-65		
Arsenic	3		7.3-117	3	Approx. Normal
BEHP	4		7.3-99	3	Not Normal
Chromium	0		7.3-119	NA	NA
Copper	0		7.3-120	NA	NA
Chlordanes	1		7.3-63	1	Normal
Dieldrin	NA	Insufficient detections	7.3-67		
DDx	2		7.3-53	2	Approx. Normal
Mercury	0		7.3-122	NA	
PAH, total	3		7.3-79	3	Approx. Normal
PCB as Aroclors	4		7.3-17	5	Normal
PCB as congeners	4		7.3-19	4	Not Normal
PCDD/F, total	1	Not evaluated by LWG	EPA revised	1	Not Normal
Tributyltin ion	NA	Insufficient data	7.3-127	NA	
Zinc	1		7.3-126	1	Normal

EPA compared the results of the statistical outlier tests with those visually observed. While there was general agreement between the two approaches, Rosner's test identified five outliers for PCBs as Aroclors, while a review of plotted data indicated the presence of only 4 distinct potential outliers. EPA also noted that two of the four outliers (UG02C and UG03B) observed for PCB Aroclors were from split samples that were also analyzed for PCB congeners, and that the congener results were 20-30 times lower than the corresponding Aroclor results. EPA also had Dr. Anita Singh (coauthor of ProUCL and Scout) conduct outlier test on the PCB Aroclor data using the more robust statistical methods available in Scout (USEPA 2009a), provided as Exhibit 5. This evaluation concluded that there were eight potential outliers; four extreme outliers and four intermediate outliers. [See Exhibit 3 Figure 7.3-17] Further, USEPA 2013b (p. 189) cautions that "it is highly likely that a data set consisting of outliers will not follow a normal distribution unless outliers are present in clusters." EPA has noted and provided several comment letters to the LWG regarding clusters as being outliers, especially the four PCB Aroclor outliers and the two DDx outliers. [Exhibit 1] As recommended in USEPA 2009a, 2013a, 2013b, the effect of outliers on the descriptive statistics was evaluated by calculating the GOF, BTV, and UCL for each contaminant with and without the outliers removed. These results are provided in Table 4.

Table 4. Results of EPA evaluation of upriver reach data set with outliers removed.

Contaminant ¹	Distribution	Units	BTV (95% UPL)	Type	CT (95% UCL)	Type
Arsenic	Normal	mg/kg	3.56	Normal-t	2.87	Normal Student's-t
BEHP	Gamma	µg/kg	103	Gamma-HW	39.8	Gamma-KM
Chlordanes	Normal	µg/kg	0.71	Normal KM-t	0.38	Normal KM-t
DDx	Normal	µg/kg	3.2	Normal KM-t	2	Normal KM-t
PAHs	Normal	µg/kg	148	Normal KM-t	73	Normal KM-t
PCB as total Aroclors (8)	Normal	µg/kg	9.08	Normal KM-t	4.67	Normal KM-t
PCB as total Aroclors (5)	Gamma	µg/kg	12.2	Gamma-HW	6	Adjusted Gamma KM
PCB as total Aroclors (4)	Lognormal	µg/kg	14.8	Lognormal KM	5.89	Normal KM-t
PCBs total congeners	Gamma	µg/kg	7.7	Gamma-HW	4.2	Gamma
total PCDD/Fs	Gamma	µg/kg	0.19	Gamma-HW	0.08	Gamma
Zinc	Normal	mg/kg	104	Normal-t	77	Normal Student's-t

1- numbers in parenthesis represent number of outliers removed

A comparison of the statistics on the full data set and with outliers removed was then conducted to evaluate the effect on the BTV and CT by removing the outliers. This comparison is provided in Table 5.

Table 5. Comparison of statistical endpoints for the full upriver reach data set and the background data with outliers removed.

Contaminant ¹	Units	Full Data Set		Outliers Removed	
		BTV (95% UPL)	CT (95% UCL)	BTV (95% UPL)	CT (95% UCL)
Arsenic	mg/kg	4	3	4	3
BEHP	µg/kg	167	52	103	40
Chlordanes	µg/kg	0.9	0.4	0.7	0.4
DDx	µg/kg	4	2	3	2
PAHs	µg/kg	353	106	148	73
PCBs as total Aroclors (8)	µg/kg	57	11	9	5
PCBs as total Aroclor (5)	µg/kg	57	11	12	6
PCB, total Aroclor (4)	µg/kg	57	11	15	6
PCB, total congener	µg/kg	21	9	8	4
PCDD/F, total	µg/kg	0.24	0.09	0.19	0.08
Zinc	mg/kg	111	79	104	77

1- numbers in parenthesis represent number of outliers removed

EPA also looked at the effect of removing successive outliers for PCBs as Aroclors, PCB as congeners, and DDx (Figures 1-3).

Figure 1. Effect of removing outliers on UPL and UCL for total PCB Aroclors.

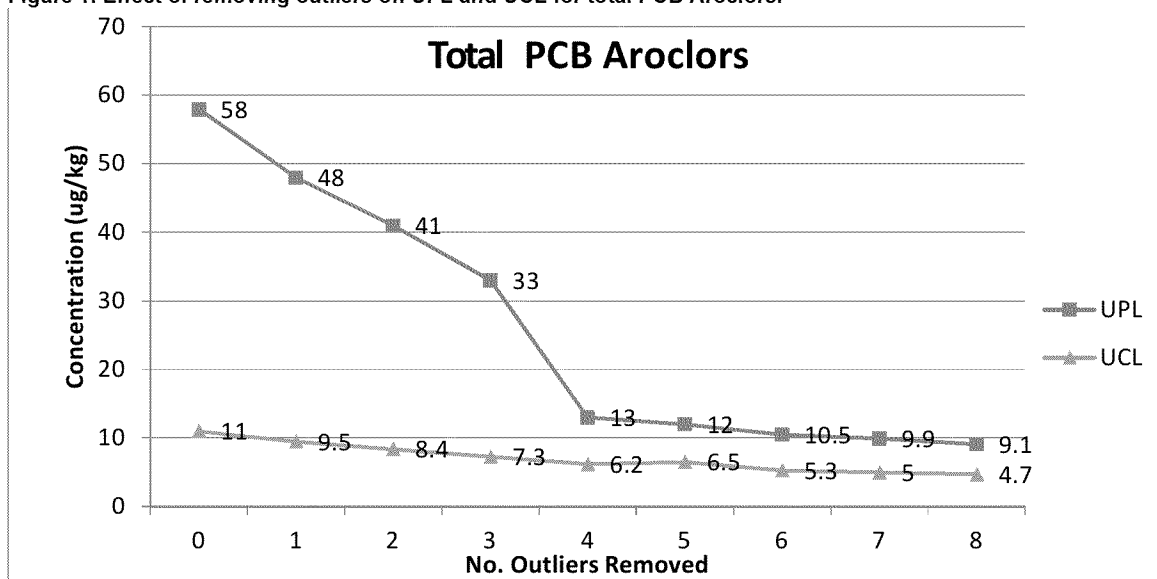


Figure 2. Effect of removing outliers on UPL and UCL for total PCB congeners.

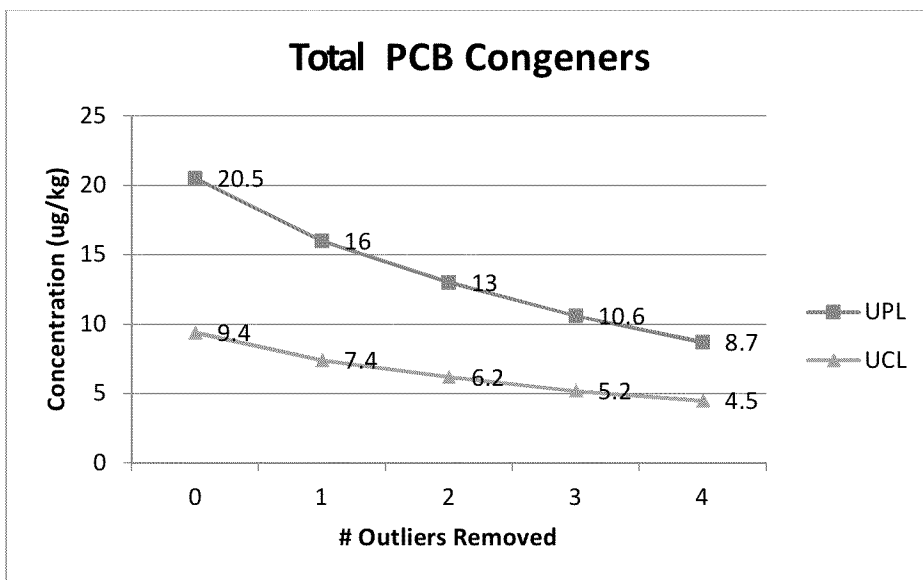
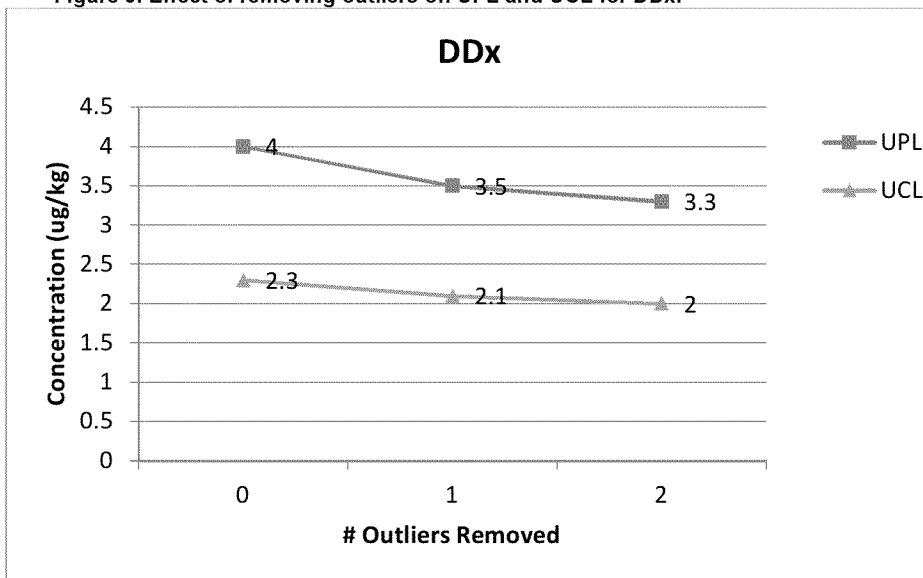


Figure 3. Effect of removing outliers on UPL and UCL for DDx.



As noted in the discussions presented in Section 7.3, not all values identified as potential outliers were ultimately removed from the final selected background data set. With the exception of PCB Aroclors, EPA removed all identified outliers consistent with the recommendations in the ProUCL Technical Guide (USEPA 2013b, Chapter 3) that the background data set should represent a single population free of outliers. EPA decided to exclude four outliers on the Aroclor data set because the fifth value identified by ProUCL was not distinctly separate from the majority of the data in the visual observations of scatter plots and Q-Q plots, and the resulting statistics were fairly comparable. The UPL and UCL for organic contaminants were then OC-corrected as described in Section 7.2.4. The final background data set developed by EPA for Section 7 of the RI is provided in Table 6.

Table 6. OC-corrected BTV and CT Background Concentrations

Contaminant	Units	BTV (95% UPL)	CT (95% UCL)
Arsenic	mg/kg	4	3
BEHP	µg/kg	155	60
Chlordanes	µg/kg	1.1	0.6
DDx	µg/kg	5	3
PAHs	µg/kg	222	110
PCB, total Aroclor	µg/kg	23	9
PCB, total congener	µg/kg	12	6
PCDD/F, total	µg/kg	0.29	0.12
Zinc	mg/kg	104	77

F. EPA'S RESPONSES TO THE LWG'S DISPUTE ISSUES

1. EPA appropriately excluded outliers based on a scientific analysis of the background data set consistent with EPA guidance.
2. EPA appropriately and scientifically calculated upriver sediment concentrations using multiple tests to identify outliers consistent with EPA guidance.
3. EPA did not arbitrarily set the number of suspected outliers to 10 for all outlier tests it performed, but rather used graphical techniques to determine the number of potential outliers for testing as recommended by EPA guidance (USEPA 2013a, 2013b).
4. EPA removed observations statistically identified as outliers based on appropriate methods consistent with EPA guidance.
5. EPA used correct statistical methods consistent with appropriate EPA guidance to evaluate Q-Q plots and other tests for outliers in datasets that contain non-detect values ("NDs").
6. EPA's justification for removing "outliers" is based on its concept that "reference area data may also contain high-biasing outliers that are either not representative of the dominant background population or are representative of specific contaminant sources." RI, Section 7.3. The LWG confuses the issue of a background reference area with recontamination potential.

Anita Singh, Ph.D. assisted EPA in the preparation of specific technical responses below through the Office of Research and Development Site Characterization and Monitoring Technical Support Center. Dr. Singh has over 30 years of experience as an academician, industrial, and environmental statistician. Dr. Singh got her MS and PhD in statistics from Purdue University, West Lafayette, Indiana in 1978. During 1980-1991, she was a professor of statistics at New Mexico Institute of Mining and Technology, New Mexico. Since 1991, she has been providing statistical guidance for environmental and chemometrics projects of the Technology Support Center (TSC) and ORD-NERL (ESD), USEPA. Dr. Singh has provided statistical guidance to EPA scientists, RPMs, and their contractors on projects dealing with various environmental media including soil, sediments, groundwater, and air. She is the technical team leader and has provided statistical guidance for over 120 CERCLA, RCRA, and unexploded ordnance (UXO) FUDS and BRAC site projects requiring assistance to address data analytic, site characterization, exposure and risk assessment, and dose-response modeling needs. Anita has taught many short courses and workshops, and given numerous technical presentations at professional meetings including conferences organized by EPA. Dr. Singh is the key developer of EPA statistical software packages: ProUCL 5.0 and Scout 2008. ProUCL is routinely used to address statistical

issues of exposure and risk assessment, groundwater monitoring, trend analysis, background evaluation, and background versus site comparison studies. These software packages are equipped with defensible statistical approaches to analyze data sets consisting of nondetect observations. She has co-authored over 60 peer-reviewed journal articles.

ISSUE #1: EPA abused its discretion by excluding outliers from the reference area data set. EPA guidance states: “One should never discard an outlier based solely on a statistical test. Instead, the decision to discard an outlier should be based on some scientific or quality assurance basis.” (EPA 2000a). EPA did not scientifically assess the outliers (e.g., weigh all available lines of evidence) to determine the reason for the elevated values, which is essential to determine whether they should be retained or removed.

The LWG provides arguments for this issue in discussions #3 and #5 of their dispute.

The LWG’s argument that the background evaluation should not eliminate outliers is not supported by extensive EPA guidance. Both USEPA 2013a and 2013b are replete with cautionary text noting that background data sets should be representative of a single environmental background population, and that “BTVs should be estimated by statistics representing the dominant background population represented by the majority of the data set. *Upper limits computed by including a few low probability high outliers (e.g., coming from the far tails of the distribution) tend to represent locations with those elevated concentrations rather than representing the main dominant background population* (EPA 2013b p-85) (emphasis in original).

As noted, “background” was intended to be representative of the recontamination potential of the broader Willamette watershed, not from the specific upstream reach defined as RMs 15.3 to 28.4. Thus, the background data set should be representative of natural and anthropogenic contaminant concentration in sediments as they are deposited in the upstream reach, and by inference, Portland Harbor itself.

The LWG’s argument that outliers should not be removed unless some basis for a likely error or discrepancy or based on other technical information or knowledge to support the decision to remove an outlier is supported by inappropriate reference to the Unified Guidance (USEPA 2009b). As stated above in [D] (p. 5), EPA does not agree that the use and application of USEPA 2009b is relevant and appropriate to evaluating background for the Portland Harbor Site.

It is not necessary for EPA to conduct a weight-of-evidence approach or to conduct an independent evaluation to remove outliers from the background data set. In the context of determining the natural and anthropogenic background from the broader Willamette watershed, the definition of an outlier is simply the classical one of not being representative of the majority of the data. The question then becomes whether the resulting calculated descriptive statistics with outliers included exhibits more the characteristics of the extreme values relative to the majority of the data. In this instance, assuming the outlier values are broadly representative of the desired

anthropogenic background population increases the likelihood of accepting a null hypothesis that the background and site populations are similar, when incorrectly rejecting the null hypothesis (Type 1 error) favors the protection of human health and the environment (EPA 2013b, p 152).

The LWG argues that data should not be discarded simply based on a statistical test because it was an error to use Rosner's or Dixon's test in the first place since those tests assume a normal distribution of the data which may not be the case.

EPA appropriately used Rosner's and Dixon's tests for the determination of outliers. EPA will address this further in EPA's response to LWG's issue #2. Further, as discussed in Section C of this response, EPA did not remove outliers based merely on a statistical test.

The LWG further argues that EPA must investigate whether or not there is any evidence to justify discarding these observations citing the Unified Guidance (USEPA 2009b). Again, EPA does not agree that USEPA 2009b is relevant and appropriate to evaluating background for the Portland Harbor Site. EPA is not discarding an observation. Rather, it has made a determination that the specific value is not representative of the desired data set. EPA disagrees that there must be definitive evidence to justify removal of an outlier.

In the LWG's discussion #5, the LWG argues that the outliers removed from the data set are more representative of the anthropogenic background than the remaining data set because the TOC of the removed outliers more closely matches the TOC in the Study Area which makes them more representative of the Study Area. They state that the average organic carbon content of all surface samples in the Study Area equals 1.79 percent [EPA note: the correct value is 1.71 percent the average organic carbon content of the reference area is 1.11 percent and the average organic carbon content of the removed outliers is 1.66 percent. They further argue that the reference area samples with lower contaminant concentrations were obtained in areas know to be higher energy than most the Study Area and exhibited lower percent fines and organic material.

As noted in Section 7 of the RI Report, although "much of the upriver reach is characterized by an exposed natural bedrock bottom and swifter currents than generally found in the study area there are pockets of reworked sand and finer-grained sediments along the margins and backwaters." The background data set was deemed "representative of the urban and suburban upland conditions of the Willamette River...uninfluenced by releases from the Portland Harbor site." With this in mind, EPA's analysis of the background data assumed those data provided by the LWG were sufficient to estimate sediment concentrations while also representing inputs from the broader watershed. The watershed has sources that cannot otherwise be reliably controlled, but EPA did not examine whether specific samples were "more representative" of the study area than others. In fact, the grain size and organic carbon content of sediments within the study area represent a broad range of values. The key factor in identifying potential outliers was not absolute concentration, or whether the sediment characteristics were "more like the study area," or "most likely to be transported." Rather, consistent with the repeated recommendations in EPA guidance, EPA's analysis focused on the degree of influence the inclusion of suspected outlier values exerted on the subsequent calculation of BTV and mean values.

In conclusion, EPA appropriately excluded outliers based on a scientific analysis of the background data set consistent with appropriate EPA guidance documents.

ISSUE #2: EPA abused its discretion (and erred scientifically) when it calculated upriver sediment concentrations using tests to identify outliers that explicitly assume a normal distribution for all populations from which the sample data were obtained.

The LWG provides arguments for this issue in discussion #1 of their dispute.

The LWG argues that EPA guidance discusses the importance of identifying whether environmental data sets are properly modeled by normal, lognormal, or some other distribution. The three EPA guidance documents cited by the LWG suggest using graphical techniques such as Q-Q plots and boxplots, as well as GOF tests and tests for outliers to determine the appropriate distribution to use, or whether to use a non-parametric method. The LWG asserts that EPA did not conduct any modeling or examining any graphical techniques in preparation of data products for section 7 of the RI Report and instead assumed a normal distribution.

The ProUCL (EPA, 2013) Technical Guide and peer reviewed articles (Singh, Singh, and Engelhardt 1997, Singh, Singh, and Iaci 2002) recommend against using a lognormal distribution on environmental data sets unless the data set is only mildly skewed. The use of a lognormal model tends to hide contamination and accommodate outliers as part of the data set, which in turn yields inflated/distorted values of UCLs, UTLs, and UPLs. Upper limits based upon data sets with outliers tend to represent the outliers rather than representing the majority of observations from the main dominant population. In environmental applications, outlier tests should be performed on raw data sets, as the cleanup decisions need to be made based upon values in the raw scale and not in log-scale or some other transformed space. Several examples are provided in Chapters 2-5 of the EPA 2013b to illustrate the deficiencies associated with the use of outlier tests on log-transformed data. (Example 2-2, Figures 2-6, 2-6a, and 2-6b of the ProUCL 5.0 Technical Guide).

In practice, it is the presence of outliers in a data set that destroys the normality of the data set; in other words, a data set consisting of outliers seldom (perhaps when only outliers are mild near the tail) can be modeled by a normal distribution. Therefore, data sets consisting of outliers often do not satisfy the normality assumption needed to use classical outlier tests (Rosner, Grubbs, and Dixon tests). The normality assumption comes into play while computing the critical values of the test statistics associated with these classical tests. It is likely that a data set without outliers can be modeled by a normal distribution. Therefore, to identify outliers based upon the Rosner test, one can use a critical value associated with the number of observations left in the data set without the number of specified/suspected outliers.

For both the Rosner and Dixon tests, it is the data set (also called the main body of the data set) obtained after removing the outliers (and not the data set with outliers) that needs to follow a normal distribution. Barnett and Lewis 1994 and Chapter 12 of EPA 2009a also state that the Rosner and Dixon outlier tests assume that the rest of the data except for the suspect outlier observation(s), are normally distributed.

USEPA 2013b recommends avoiding the use of a lognormal model, as its use tends to accommodate outliers. Even the use of graphical methods (e.g., Q-Q plot) fails to identify outliers in the log-scale. USEPA (2009a) also states that the data set without the outliers should follow a normal distribution.

The LWG contends that EPA's use of outlier tests on raw (untransformed) data is "arguably consistent" with the statement in USEPA 2013a that outlier tests should be performed on raw

data, but is “contradictory to the guidance as a whole.” A more complete reading of EPA guidance on determining background paints a completely different picture. Section 4.3 of EPA’s Guidance for Comparing Background and Chemical Concentrations in Soil for CERCLA Sites (USEPA, 2002a) notes that “Outliers in the site data set have different implications from outliers in the background data set. For example, an onsite outlier can indicate a hot spot, which indicates that the one spot needs attention. An outlier in the background data set, however, might indicate that one of the background samples was collected in a location that is not truly background. In such a case, an outlier test should be used (along with a qualitative study of where the sample in question was collected) to see if that data point should be discarded from the background set.” USEPA 2013b notes the following:

“It is implicitly assumed that the background data set used to estimate BTVs represents a single statistical population. However, since outliers (well-separated from the main dominant data) are inevitable in most environmental applications, some outliers such as the observations coming from populations other than the background population may also be present in a background data set. Outliers, when present, distort decision statistics of interest (e.g., upper prediction limits [UPLs], upper tolerance limits [UTLs])...” (p.84).

“The BTVs should be estimated by statistics representing the dominant background population represented by the majority of the data set. *Upper limits computed by including a few low probability high outliers (e.g., coming from the far tails of data distribution) tend to represent locations with those elevated concentrations rather than representing the main dominant background population*” (emphasis in original, p. 85).

“Established background data set: represents background conditions free of outliers which potentially represent locations impacted by the site and/or other activities. An established background data set should be representative of a single environmental background population. This can be determined by using a normal Q-Q plot on a background data set. If there are no jumps and breaks in the normal Q-Q plot, the data set may be considered to represent a single environmental population. Outliers when present in a data set result in inflated values of the various decision statistics including: UPL, UTL, and USL. The use of inflated statistics as BTV estimates tends to result in a higher number of false negatives” (p. 87).

“The inclusion of outliers in a background data set tends to yield distorted (inflated) estimates of BTVs. Outlying observations which are significantly higher than the majority of the background data may not be used in establishing background data sets and in the computation of BTV estimates” (p. 89).

“When high outlying observations are present in a background data set, the higher order statistics may represent observations coming from the contaminated onsite/offsite areas. Decisions made based upon outlying observations or distorted upper limits can be incorrect and misleading” (p. 89).

“The decision limits and test statistics should be computed based upon the majority of data representing the main dominant population and not by accommodating a few low probability outliers resulting in distorted and inflated values of the decision statistics. The inflated statistics tend to represent the locations with those elevated observations rather than representing the main dominant population” (pp. 89-90).

Hence, while it may be possible to fit the background data to various distributions, EPA's background guidance is consistent and clear that background data sets should represent the majority of the data while excluding outliers.

In conclusion, EPA appropriately and scientifically calculated upriver sediment concentrations using tests to identify outliers consistent with appropriate EPA guidance documents. EPA did not assume a normal distribution for the data, but used the most appropriate methods available in the literature to perform outlier analysis, GOF tests, and compute UCLs and BTVs.

ISSUE #3: EPA abused its discretion by arbitrarily setting the number of suspected outliers to 10 for all outlier tests it performed, contrary to the advice of EPA guidance documents, which recommend using graphical techniques to determine the number of potential outliers for testing.

The LWG provides arguments for this issue in discussion #2 of their dispute.

The LWG argues that determination of an "outlier" will depend on the underlying distribution and the model used to interpret the data and if the decision is made that the distribution is other than the normal distribution, then it does not make sense to use Dixon's or Rosner's test on the raw data. The LWG contends that EPA did not provide any evidence that it used the graphical presentations to aid in identifying the number of potential outliers. The LWG provides two figures to support their arguments: Figure 1 shows the distribution of the number of outliers indicated by Rosner's test for various probability distributions on a made-up data set and Figure 2 provides GOF Q-Q plots using various transformations of the censored data (i.e., data excluding outliers).

It is not clear to EPA what "evidence" would be required to support its contention that observation of Q-Q and scatter plots was used in conjunction with outlier tests. No evidence beyond a similar statement was provided by the LWG in its draft final version of Section 7. There is nothing wrong in performing Rosner's tests for 10 suspected outliers. The test will not identify 10 outliers if there are less than 10 present in the data set. However, it is desirable to use graphical displays to determine the relative extremeness/ clustering of the identified outliers. The graphical displays also give some idea about the data distribution. The normality assumption is not required to perform the outlier tests and outlier tests should be performed in the original raw scale. The Rosner's test can be used for up to 10 suspected outliers. However, once the outliers have been identified, they should be investigated further to identify the reasons of their occurrences. The use of graphical displays is suggested to determine the relative extremeness/clustering of the identified outliers.

EPA performed outlier tests correctly for its revisions of Section 7. As already noted, there is no requirement that the data has to follow a normal distribution to be able to use Rosner's or Dixon's tests. Rosner's test can be used for up to 10 suspected outliers, and there is nothing wrong with using the test to determine the number of potential outliers. However, once potential outliers have been identified, they should be investigated further.

As illustrated by the following discussion of the evaluation of the Aroclor data, it is clear that EPA did not simply rely on the results of the outlier tests. EPA used exploratory Q-Q plots,

reviewed the scatter plots, and evaluated the overall effect potential outliers exhibited on descriptive statistics before making a decision to include or exclude suspected outliers.

A Q-Q plot for the Aroclor data set with all 48 observations is shown on Figure 4, and a Q-Q plot of the 25 detected observations only is provided on Figure 5 to provide information about data distribution and presence of outliers.

Figure 4. Q-Q plot of Aroclors displaying all 48 data points

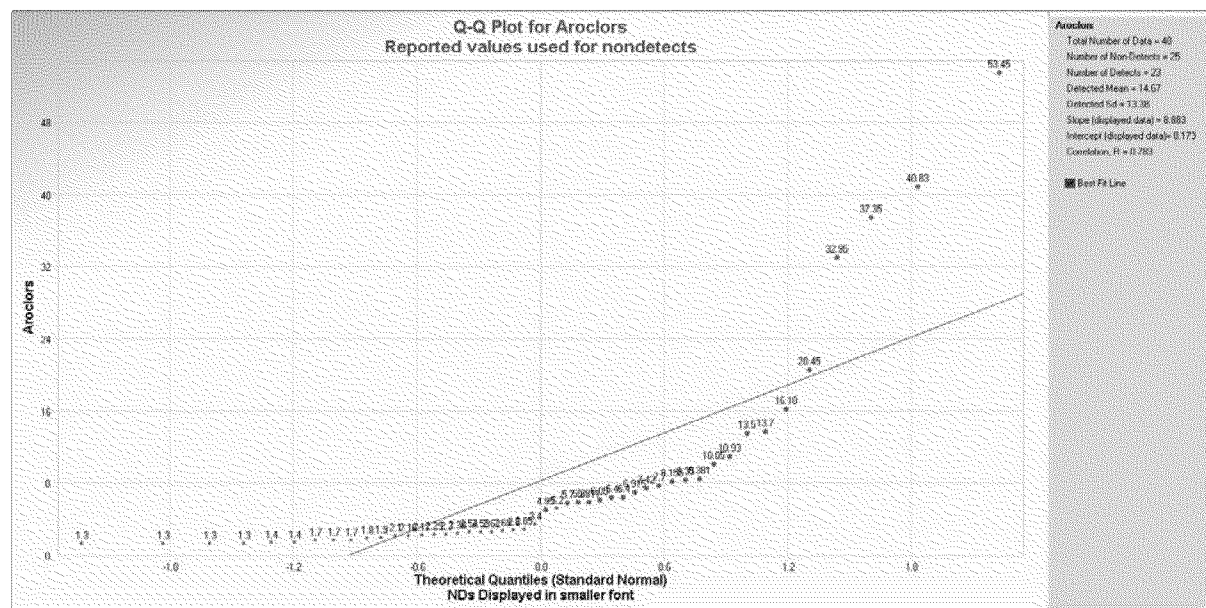
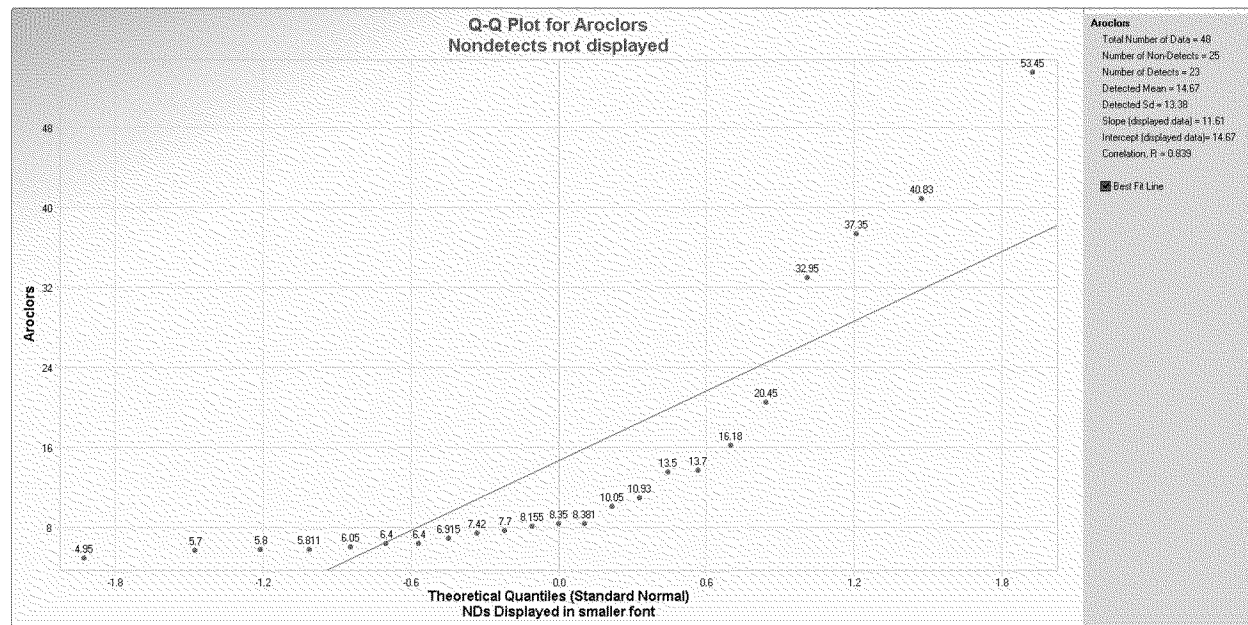


Figure 5. Q-Q plot of detected Aroclors displaying 25 detected data points



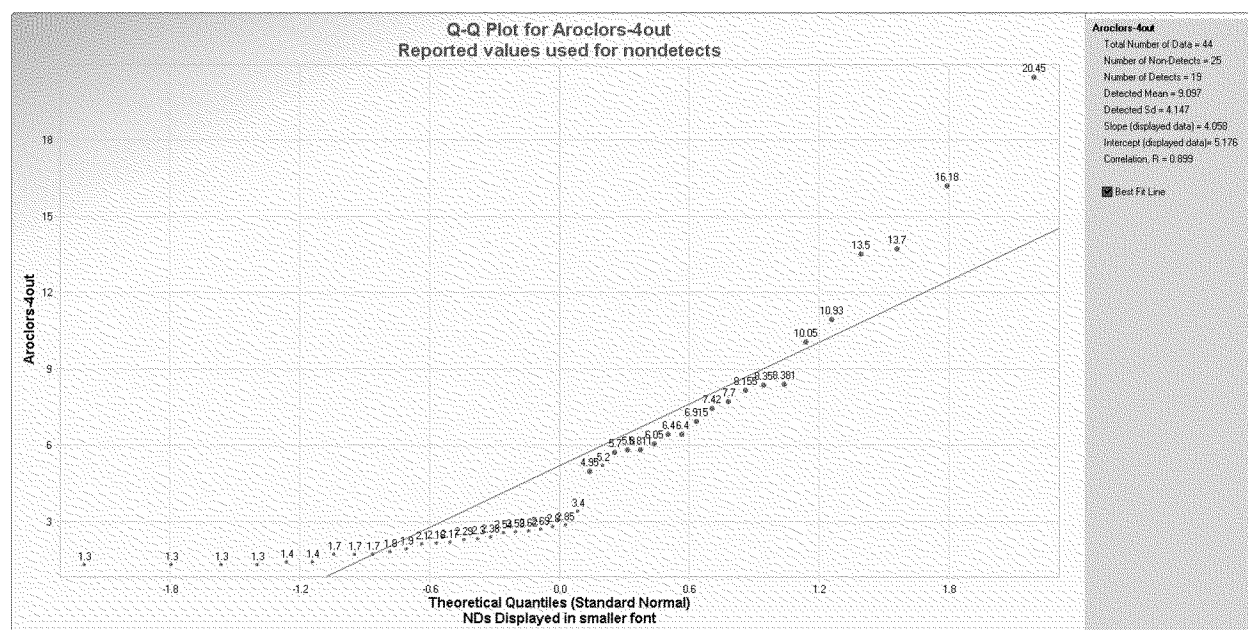
As can be seen in these figures, there are at least 4 detected observations which are well separated from the rest of the data set and potentially represent outliers with respect to the background data set. EPA correctly used Rosner's test on the Aroclor data set, which identified 5 outliers at a 0.05 level of significance and 4 outliers at a 0.01 level of significance, as shown in Figure 6 below.

Figure 6. ProUCL 5.0 output file for Rosner's outlier test on the PCB Aroclor data set.

Rosner's Outlier Test for 10 Outliers in PCBs								
Total N		48						
Number NDs		25						
Number Detects		48						
Mean with NDs=DL/2		7.601						
SD with NDs=DL/2		11.44						
Number of data		48						
Number of suspected outliers		10						
NDs replaced with half value.								
#	Mean	sd	Potential outlier	Obs. Number	Test value	Critical value (5%)	Critical value (1%)	
1	7.601	11.32	53.45	10	4.051	3.11	3.46	
2	6.626	9.326	40.83	7	3.668	3.1	3.46	
3	5.882	7.896	37.35	21	3.985	3.09	3.45	
4	5.183	6.384	32.95	20	4.35	3.09	3.44	
5	4.552	4.834	20.45	9	3.289	3.08	3.43	
6	4.182	4.215	16.18	39	2.846	3.07	3.418	
7	3.897	3.821	13.7	44	2.565	3.06	3.406	
8	3.657	3.537	13.5	11	2.783	3.05	3.394	
9	3.411	3.207	10.93	38	2.345	3.04	3.382	
10	3.219	3.004	10.05	43	2.274	3.03	3.37	
For 5% significance level, there are 5 Potential Outliers								
53.45, 40.83, 37.35, 32.95, 20.45								
For 1% Significance Level, there are 4 Potential Outliers								
53.45, 40.83, 37.35, 32.95								

The four extreme outliers represent observations coming from a population different from the population represented by the remaining 44 observations as confirmed from the following graph shown on Figure 7.

Figure 7. Q-Q plot of 44 observations without the 4 extreme outliers shown in Figure 4.



Except for the result of 20.45 $\mu\text{g/kg}$, the data represented in Figure 7 appear to be coming from a single population.

To determine the effect of this value, BTVs and UCLs were computed without the 4 outliers and without the 5th potential outlier. The differences between these calculated values are minor, suggesting that the intermediate potential outlier of 20.45 $\mu\text{g/kg}$ may be considered as coming from the population represented by the data set consisting of 43 observations.

EPA acknowledges that the information provided to the LWG on December 6, 2013 incorrectly notes five outliers in Section 7.3.11 and the attached Excel file. In our discussions with LWG regarding the Aroclor data, it was agreed by all parties that the result from Sample UG02A did not appear distinct from the rest of the data, and that it would be included as part of the background data set. The resulting OC-corrected 95 percent UPL and UCL are 23 $\mu\text{g/kg}$ and 9 $\mu\text{g/kg}$, respectively. [see Table 6]

In conclusion, EPA did not arbitrarily set the number of suspected outliers to 10 for all outlier tests it performed and used graphical techniques to determine the number of potential outliers for testing as recommended by EPA guidance.

ISSUE #4: Contrary to EPA guidance documents, EPA discarded observations statistically identified as outliers based on an improperly applied statistical test without investigating whether any evidence justified discarding those observations, such as analytical quality issues or site-specific environmental conditions.

The LWG provides arguments for this issue in discussion #3 of their dispute.

The LWG provides no evidence that EPA did not consider whether any evidence justified discarding the observations, and such a claim is not supported by the evidence. For example, The

LWG asserts that “EPA did not provide any evidence that it used the graphical presentations to aid in identifying the number of potential outliers.” See [E] above and Exhibit 3 where EPA describes how it evaluated and reviewed the graphical presentations provided by the LWG. In addition, EPA provided to the LWG on November 5, 2013 as Figure 7.2-2 [see Exhibit 6] a scatter plot of paired Aroclor-congener results for the upstream reach for samples that were analyzed by both EPA Methods 8082 and 1668a. EPA noted that the correlation between Aroclor and congener results was particularly poor in this reach. The reported results for samples UG02C were 16.7 and 2.7 µg/kg, respectively, and 17.1 and 1 µg/kg, respectively, for sample UG03B. Thus, there was every reason to suspect that the Aroclor data were in fact not representative and should be excluded as outliers. The LWG’s assertion that EPA conducted no evaluation regarding whether specific outliers should be excluded is thus unfounded.

In conclusion, EPA removed observations statistically identified as outliers based on appropriate methods consistent with EPA guidance.

ISSUE #5: EPA failed to use correct statistical methods to evaluate Q-Q plots or to otherwise formally test for outliers in datasets that contain nondetect values (“NDs”), such as the use of Tobit regression.

The LWG provides arguments for this issue in discussion #4 of their dispute.

The LWG makes three arguments:

- a. The LWG argues that EPA incorrectly developed Q-Q plots for data that contains nondetects because EPA guidance (USEPA 2009a; USEPA 2013a, b) makes incorrect recommendations for the treatment of NDs in constructing Q-Q plots.
- b. The LWG argues that EPA incorrectly conducted GOF tests for data that contains nondetect values because EPA guidance (USEPA 2009a; USEPA 2013a, b) deviates from standard statistical approaches.
- c. The LWG argues that EPA incorrectly determined outliers on data that contain nondetect values because EPA guidance (USEPA 2009a; USEPA 2013a, b) incorrectly recommends treatment of NDs when performing either Dixon’s or Rosner’s test.

In response to [a], EPA disagrees with the LWG’s statement that ProUCL computes incorrect plotting positions for data sets consisting of NDs when generating Q-Q plots. It is only detected outlying observations may require additional investigation for the purpose of identifying potential outliers on the right tail of the distribution. Therefore, from an exploratory point of view, there is no need to retain place holders by computing plotting positions used to impute NDs, as the objective is not to impute NDs (USEPA 2013b, p. 33). And since the LWG argues that constructing censored Q-Q plots is “simply throwing away data that one does not know what to do with and is patently wrong,” it isn’t clear why LWG Figures 2.1 – 2.33 purport to show correct Q-Q plots, when as clearly noted in the figures, they are censored Q-Q plots.

In response to [b], the LWG misrepresents methods described in USEPA guidance (2009a, 2013a, and 2013b) for GOF tests on data sets containing NDs. None of the chapters of the ProUCL Guidance (USEPA 2013a and 2013b) and modules of the ProUCL software recommend the use of substitution methods. Historically, for data sets with NDs, environmental scientists

(Gilbert 1987, Gibbons 1994, Millard and Neerchal 2002, USEPA 2006a, 2006b) have been using/suggesting the use of various parametric methods (Cohen's maximum likelihood method [MLE], delta method) for data sets with a single detection limit (DL); substitution methods (setting NDs to half of their respective DLs or to their respective DLs); or regression on order statistics (ROS) methods (Helsel 2005 Millard and Neerchal 2002). For data sets consisting of NDs with multiple DLs, the DL/2 method has been most commonly used until more rigorous methods (e.g., Kaplan-Meier [KM] method, bootstrap methods) became available. Based upon the findings of the research conducted by National Exposure Research Laboratory scientists and developers of ProUCL and Scout software packages (Singh and Nocerino 2002, Singh, Maichle, and Lee 2006), throughout ProUCL and Scout packages, the developers have placed emphasis on not using substitution methods for computation of UCLs, UTLs, UPLs, GOF tests, and t-tests. The availability of the historical methods in the ProUCL 5.0 and Scout 2008 packages should not be interpreted as recommended or endorsed methods to compute UCLs on means, UTLs, UPLs, upper percentiles, and GOF test statistics. It is incorrect to infer that ProUCL 5.0 recommends use of the DL/2 method just because the option is available. Several sections of the ProUCL 5.0 document specifically recommend not to use the DL/2 method (Sections 1.11.1 and 4.2.3). For data sets with NDs, all UCL and BTV outputs generated by ProUCL 5.0 include the following message:

“DL/2 is not a recommended method, provided for comparisons and historical reasons”

The methods to compute rigorous upper limits for skewed data sets with NDs which are described in Helsel 2005, 2012, Millard and Neerchal 2002, USEPA 2009b do not properly adjust for skewness. The use of the upper limits computation methods described in Helsel 2005, 2012, and Millard and Neerchal 2002, tends to underestimate the environmental parameters of interest including UCLs and BTVs.

ProUCL 5.0 provides several options to perform GOF tests for data sets consisting of NDs. These options include excluding all NDs, replacing NDs with DL/2, and ROS methods. The GOF test obtained using the DL/2 method is not used in any decision making process and is retained in ProUCL 5.0 for historical and comparison purposes. The ROS method is available in ProUCL as it has been a commonly used method (e.g., Helsel 2005) and has been cited in several documents, including USEPA 2009b. For ROS methods, one assumes that the entire data set follows a certain distribution and NDs are imputed using the assumed distribution. Quantiles associated with ROS methods are computed using their respective plotting positions (Singh, Maichle, and Lee 2006).

In response to [c], the key purpose of performing outlier tests on an environmental data set is to identify those outlying observations. Nondetect observations represent locations where the analyte of interest has not been detected and identification of outlying nondetect values is not required. As described on p 192 of USEPA 2013b, substitution methods (DL/2 method and DL method) are included for exploratory purposes. There is nothing wrong in looking at a data set in several different ways. This practice of using exploratory as well as statistical methods is commonly used by scientists and researchers. In EPA's review of the data, outliers identified in the statistical tests corresponded with outliers identified through visual observation of the data.

Therefore, EPA has confidence that the statistical determination of outliers did not produce questionable result.

In conclusion, EPA used correct statistical methods consistent with appropriate EPA guidance to evaluate Q-Q plots and other tests for outliers in datasets that contain non-detect values (NDs).

ISSUE #6: EPA’s justification for removing “outliers” is based on its concept that “reference area data may also contain high-biasing outliers that are either not representative of the dominant background population or are representative of specific contaminant sources.” RI, Section 7.3. However, because upstream bedded sediments with elevated concentrations may be transported downstream to the Site, it is important for the reference area data to represent the total reference area population, not a post-hoc background population constructed by the removal of valid data.

The LWG provides arguments for this issue on pages 1-4 of their dispute.

The LWG argues that the background data represents an estimate of mobile upstream contaminant mass and a gross underestimation of this mass is likely to result in a site remedy with unachievable remedial action goal. The LWG provides an example of a flood that scours one foot of sediment from the entire upstream reach, homogenizes the sediment in the water column, and then transports the sediment downstream to be deposited in the study area. The LWG contends that the only way to accurately estimate this mobile homogenized concentration is to estimate the mass, regardless of location and concentration. They further argue that this is the very reason that all the data must be used to accurately estimate this mass. The LWG then goes on to describe the method for evaluating this mass as developing a surface weighted average concentration (SWAC) for the background area and that excluding outliers in the development of a SWAC would provide critical errors in the evaluation of whether site-specific releases in the study area have resulted in elevated chemical concentrations in the Study Area, and achievable concentrations for remedial alternative in the FS Report.

To be clear, data collected for the RI clearly show the contamination released from site sources is far greater than concentrations associated solely with non-site related sources. Thus, the calculation of background is most relevant to assessing appropriate long-term cleanup goals for the site.¹² As EPA has stated before, the background data set is meant to represent the loading from the broader watershed, not from the upriver reach of the river. However, the LWG has conducted no analysis of the potential sediment and contaminant mass available in this reach of

¹² EPA guidance (USEPA 2002a, USEPA 2013a, USEPA 2013b) describes appropriate statistical methods to evaluate whether site-specific releases in the study area have resulted in elevated chemical concentrations. The LWG did not conduct such an analysis in the 2011 draft RI Report. The LWG now contends that the appropriate statistical evaluation is a SWAC is not supported by any EPA guidance and the data were not collected with sufficient density to even consider using such an approach. Further, specific reaches were not sampled in the upriver reach (RM 19 to RM 21) because of known sources.

the river for scour and re-deposition, as all mass-loading analyses in the 2011 draft RI and 2013 draft FS included data collected down to RM 11. Consequently, even assuming the flood scenario posited on page 2 of the LWG's dispute submittal, the representativeness of outlier values in accurately representing the available contaminant mass is questionable. Thus, in this context, EPA is justified in excluding the outlier values from the analysis.

EPA acknowledges that there are sources in the upriver reach and the downtown corridor that will affect the ability of the Study Area to equilibrate to the background concentrations. However, this is not a justified reason for concluding that the background concentrations or the long-term goals should be set to the current loadings. The FS will evaluate this appropriately in the evaluation of long-term effectiveness and will discuss the uncertainty in the ability of the site to reach the background levels in the long-term. In establishing final remediation goals for the proposed plan, EPA will select values that are achievable using evaluations in the FS consistent with EPA guidance.

In conclusion, EPA's justification for removing "outliers" is based on the concept that "reference area data may also contain high-biasing outliers that are either not representative of the dominant background population or are representative of specific contaminant sources." RI, Section 7.3. The LWG argument confuses the issue of a background reference area with recontamination potential. These are two, distinctly different issues.

References:

- Gilbert, R.O. 1987. Statistical Methods for Environmental Pollution Monitoring. Van Nostrand Reinhold, New York.
- Gibbons, R. 1994. Statistical Methods for Groundwater Modeling. John Wiley and Sons, NY.
- Helsel, D.R. 2005. Nondetects and Data Analysis. Statistics for Censored Environmental Data. John Wiley and Sons, NY.
- Helsel, D.R. (2012). Statistics for Censored Environmental Data Using Minitab and R. Second Edition. John Wiley & Sons, Hoboken, New Jersey.
- Millard, S.P., and N.K. Neerchal. (2002). Environmental Statistics with S-PLUS. CRC Press, Boca Raton, Florida.
- Singh, A. and Nocerino, J.M. 2002. Robust Estimation of the Mean and Variance Using Environmental Data Sets with Below Detection Limit Observations, Vol. 60, pp 69-86
- Singh, A., Maichle, R., and Lee, S. 2006. *On the Computation of a 95% Upper Confidence Limit of the Unknown Population Mean Based Upon Data Sets with Below Detection Limit Observations*. EPA/600/R-06/022, March 2006.
<http://www.epa.gov/osp/hstl/tsc/softwaredocs.htm>
- USEPA. 1995. Determination of Background Concentrations of Inorganics in Soils and Sediments at Hazardous Waste Sites. U.S. Environmental Protection Agency, Office of Solid Waste and Emergency Response, Washington, DC.
- USEPA. 2002a. Guidance for Comparing Background and Chemical Concentrations in Soil for CERCLA Sites. EPA 540-R-01-003, September, 2002. Office of Emergency and Remedial Response, U.S. Environmental Protection Agency, Washington, D.C.
- USEPA. 2002b. Role of Background in the CERCLA Cleanup Program. OSWER 9285.6-07P, April, 2002. Office of Solid Waste and Emergency Response, Office of Emergency and Remedial Response, U.S. Environmental Protection Agency, Washington, D.C.
- USEPA. 2005. Contaminated Sediment Remediation Guidance for Hazardous Waste Sites. EPA-540-R-05-012. U.S. Environmental Protection Agency, Office of Solid Waste and Emergency Response, Washington, DC. December, 2005.
- USEPA. (2006a). *Data Quality Assessment: A Reviewer's Guide*. EPA QA/G-9R. EPA/240/B-06/002, February 2006. Office of Environmental Information, U.S. Environmental Protection Agency, Washington, D.C.

USEPA. (2006b). *Data Quality Assessment: Statistical Methods for Practitioners*. EPA QA/G-9S. EPA/240/B-06/003, February 2006. Office of Environmental Information, U.S. Environmental Protection Agency, Washington, D.C.

USEPA. 2009a. Scout 2008 Version 1.0 User Guide (Second Edition, December 2008). EPA/600/R-08/038, February 2009. Office of Research and Development, National Exposure Research Laboratory, Environmental Sciences Division, Technology Support Center, Characterization and Monitoring Branch, U.S. Environmental Protection Agency, Las Vegas, Nevada.

USEPA. 2009b. Statistical Analysis of Groundwater Monitoring Data at RCRA Facilities: Unified Guidance. EPA 530-R-09-007, March 2009. Office of Resource Conservation and Recovery, Program Implementation and Information Division, U.S. Environmental Protection Agency, Washington, D.C.

USEPA. 2013a. ProUCL Version 5.0.00 Users Guide. EPA/600/R-07/041, September 2013. Office of Research and Development, U.S. Environmental Protection Agency, Washington, D.C.

USEPA. 2013b. ProUCL Version 5.0.00 Technical Guide. EPA/600/R-07/041, September 2013. Office of Research and Development, U.S. Environmental Protection Agency, Washington, D.C.

Exhibits:

Exhibit 1 – Compendium of comments, technical memoranda, and discussions on background.

Exhibit 2 – EPA’s final edits to Section 7 of the RI.

Exhibit 3 – LWG’s figures from 2011 draft RI with EPA’s visual and observational markings

Exhibit 4 – EPA’s working files (Excel Spreadsheet) for evaluation of the background data set.

Exhibit 5 – Classical and Robust Outlier Tests on Total PCB values-R10

Exhibit 6 – Figure 7.2-2 Scatter plot of paired Aroclor-congener results for the upstream reach for samples that were analyzed by both EPA Methods 8082 and 1668a